

LETTER

 Communicated by Jaap van Pelt

The Combinatorics of Neurite Self-Avoidance

Elizabeth M. Forbes

elizabeth.forbes@uqconnect.edu.au

*Queensland Brain Institute, University of Queensland, St. Lucia,
QLD 4072, Australia*

Jonathan J. Hunt

jjh@42quarks.com

Geoffrey J. Goodhill

g.goodhill@uq.edu.au

*Queensland Brain Institute and School of Mathematics and Physics,
University of Queensland, St. Lucia, QLD 4072, Australia*

During neural development in *Drosophila*, the ability of neurite branches to recognize whether they are from the same or different neurons depends crucially on the molecule Dscam1. In particular, this recognition depends on the stochastic acquisition of a unique combination of Dscam1 isoforms out of a large set of possible isoforms. To properly interpret these findings, it is crucial to understand the combinatorics involved, which has previously been attempted only using stochastic simulations for some specific parameter combinations. Here we present closed-form solutions for the general case. These reveal the relationships among the key variables and how these constrain possible biological scenarios.

1 Introduction ---

The correct wiring of the nervous system during neural development requires growing axons and dendrites (collectively termed *neurites*) to target appropriate locations with exquisite sensitivity. In the past two decades, it has become clear that this targeting is choreographed by a relatively small number (order 10^2) of distinct molecular guidance cues (Dickson, 2002; Huber, Kolodkin, Ginty, & Cloutier, 2003; Chilton, 2006). However, understanding how these cues actually operate to achieve correct wiring remains an extremely active area of experimental research (Mortimer, Fothergill, Pujic, Richards, & Goodhill, 2008; O'Donnell, Chance, & Bashaw, 2009). One crucial problem in the formation of appropriate wiring patterns in many systems is neurite self-avoidance. This refers to the ability of neurite

Elizabeth Forbes and Jonathan Hunt contributed equally to this article.

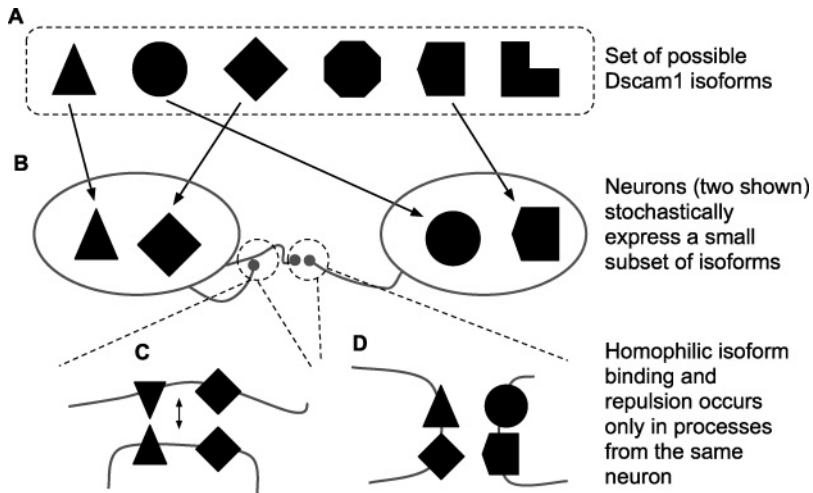


Figure 1: Schematic showing how *Dscam1* is believed to facilitate neuronal self-avoidance. (A) There are a large number ($i \approx 20,000$) of possible *Dscam1* isoforms. (B) Each neuron stochastically expresses a small subset of the isoforms ($m \approx 30$ isoforms). (C) When processes from the same neuron encounter one another, the isoforms homophilically bind. This binding generates repulsion. (D) However, when processes from different neurons encounter one another, they are unlikely to be expressing the same isoforms, so homophilic binding does not occur and the processes do not repel.

branches to recognize whether they are from the same or different neurons, so that branches from the same neuron may avoid contacting each other. It allows neurons to make a large number of connections to other neurons over a wide area while avoiding wasting resources on recursive connections. Given the small number of guidance cues available relative to the total number of neurons in even a small nervous system such as that of the fly, how could each neuron be labeled with a separate molecular identity to allow such self-avoidance?

Remarkably, recent work has identified a pivotal role for the *Dscam* family of immunoglobulin cell surface proteins in this process (reviewed in Hattori, Millard, Wojtowicz, & Zipursky, 2008). In particular in *Drosophila*, alternative splicing of the single gene *Dscam1* can generate 19,008 protein isoforms. Individually these exhibit isoform-specific binding, and homophilic recognition results in repulsion. Individual neurons express a unique set of *Dscam1* isoforms. When their neurites contact, they undergo homophilic binding, which generates repulsion, whereas when other neurons, which have a different set of isoforms, are encountered, homophilic binding and repulsion do not occur (see Figure 1). The number of *Dscam1*

isoforms expressed by each neuron is believed to be in the range 10 to 50, and it has been proposed that this set of isoforms is chosen stochastically from the set of all possible isoforms (Hattori et al., 2009). Although little is known biologically about how this might occur, understanding the consequences of this assumption reduces to a problem in combinatorics.

However, a simple calculation reveals a problem: even with a large number of isoforms to choose from, neurons randomly expressing Dscam1 isoforms are likely to encounter other neurons with at least one isoform in common, potentially leading to inappropriate non-self-avoidance. This is analogous to the counterintuitively large probability that two children in a class at school share the same birthday. Hattori et al. (2009) therefore proposed that some degree of “sharing” is allowed: neurites can share some isoforms and not be repelled, provided the proportion of isoforms in common is small compared to the total number of isoforms expressed. The lack of repulsion when some of the isoforms are shared may be because a threshold amount of isoform binding is required to generate repulsion.

The combinatorial implications of this assumption are nontrivial to calculate. Hattori et al. (2009) addressed this numerically by performing stochastic simulations for some specific situations (an example is shown in Table 1) but did not attempt the general case. Here we first analytically derive a relatively simple approximate closed-form solution and explore the relationships this implies between the key variables involved. Second, we show that this problem is analogous to a type of “collision” problem recently addressed in the literature of combinatorics, for which an exact solution has been derived. Numerical evaluations of these formulas show that our approximate solution agrees very closely with the exact solution and that these also agree closely with the stochastic simulations of Hattori et al. (2009). Together these results deepen our understanding of neurite self-avoidance and may be a useful guide for interpreting biological data in the future. These calculations also provide a general approach for addressing similar combinatorial problems in neural development.

2 Derivation of an Approximate Solution

We begin with a pairwise comparison of two neurons and determine the probability that these neurons share exactly j isoforms. We use m as the number of isoforms expressed per neuron and i as the total number of possible isoforms (the terminology is summarized in Table 2). We assume that the order in which the isoforms are selected does not matter.

First, consider the problem with the simplifying assumption that a neuron expresses m distinct isoforms (the same isoform cannot be chosen more than once per neuron). Now consider the case where the two neurons do not share any isoforms. In this situation, let the first neuron choose m isoforms. If the second neuron could choose isoforms without restriction, then the number of ways it could choose m isoforms is $\binom{i}{m}$. In order to have different

Table 1: Probability That a Pair of Neurons, Each Expressing 30 Isoforms Randomly Selected from a Set of m Isoforms, Has k or More Isoforms in Common as Generated by Monte Carlo Simulations in Hattori et al. (2009).

k	Number of Isoforms (i)									
	20,000	10,000	5000	2000	1000	500	200	100		
1	0.044	0.086	0.16	0.36	0.59	0.83	0.99	1.0		
2	0.00092	0.0036	0.013	0.071	0.22	0.52	0.93	1.0		
3	1.2e-05	9.3e-05	0.00070	0.0091	0.053	0.24	0.79	0.99		
4	1.2e-07	1.7e-06	2.5e-05	0.00082	0.0095	0.082	0.57	0.96		
5	0	2.2e-08	6.7e-07	5.5e-05	0.0013	0.022	0.35	0.88		
6	0	0	1.0e-08	3.0e-06	0.00013	0.0045	0.18	0.75		
7	0	0	0	1.3e-07	1.1e-05	0.00076	0.074	0.57		
8	0	0	0	4.0e-09	7.2e-07	0.00010	0.026	0.38		
9	0	0	0	0	4.0e-08	1.1e-05	0.0073	0.22		
10	0	0	0	0	1.0e-09	9.7e-07	0.0017	0.11		
11	0	0	0	0	0	7.4e-08	0.00034	0.044		
12	0	0	0	0	0	5.0e-09	5.6e-05	0.016		
13	0	0	0	0	0	0	7.7e-06	0.0046		
14	0	0	0	0	0	0	8.6e-07	0.0011		
15	0	0	0	0	0	0	8.5e-08	0.00023		
16	0	0	0	0	0	0	3.0e-09	4.0e-05		
17	0	0	0	0	0	0	0	5.6e-06		
18	0	0	0	0	0	0	0	6.3e-07		
19	0	0	0	0	0	0	0	5.3e-08		
20	0	0	0	0	0	0	0	5.0e-09		
21	0	0	0	0	0	0	0	0		

Note: Hattori et al. (2009) used 10^6 iterations for their simulation so events that occur with probability $\lesssim 10^{-9}$ were approximated as occurring with a probability of 0.

Table 2: Notation Used Throughout.

n	Number of neurons
N	Number of neuron pairs
i	Total number of isoforms
m	Number of isoforms expressed per neuron
$C(j)$	Probability that a pair of neurons shares <i>exactly</i> j isoforms
$Q(k)$	Probability that a pair of neurons shares <i>at least</i> k isoforms
P	Probability that all neurons in a population are distinct from each other
r	Number of nondistinct isoforms on a single neuron
$\binom{p}{q}$	Choosing q from p without repeats
$\left(\binom{p}{q}\right)$	Choosing q from p with repeats

Note: This notation is similar to Hattori et al. (2009).

isoforms from the first isoform, the second neuron must choose m isoforms from $i - m$ isoforms, that is, $\binom{i-m}{m}$. Thus, the probability that two neurons do not share any isoforms is

$$C(0) = \frac{\binom{i-m}{m}}{\binom{i}{m}}. \quad (2.1)$$

Later we relax the assumption that all m isoforms chosen by each neurons are distinct. In the case where one isoform is the same, there are m ways to choose an isoform from the first neuron that will also be expressed on the second neuron such that the two neurons have a single isoform in common. The second neuron can choose only $m - 1$ isoforms, as the one in common has already been chosen. In addition, as the isoform in common has been chosen, there are now $i - m + 1$ isoforms from which to choose. The probability of having one isoform in common is therefore

$$C(1) = m \frac{\binom{i-(m-1)}{m-1}}{\binom{i}{m}} \quad (2.2)$$

$$= m \frac{\binom{i-m+1}{m-1}}{\binom{i}{m}}. \quad (2.3)$$

The case of two isoforms in common is similar, except that there are now $\binom{m}{2}$ ways to choose these two isoforms, and the choice is $m - 2$ isoforms

from $i - m + 2$ isoforms. Extending the argument to the case of exactly j isoforms in common yields a probability of this occurring of

$$C(j) = \binom{m}{j} \frac{\binom{i-m+j}{m-j}}{\binom{i}{m}}. \tag{2.4}$$

Therefore, the probability that two neurons share fewer than k isoforms is

$$\sum_{j=0}^{k-1} C(j) = \sum_{j=0}^{k-1} \binom{m}{j} \frac{\binom{i-m+j}{m-j}}{\binom{i}{m}}, \tag{2.5}$$

and the probability, $Q(k)$, of two neurons sharing at least k isoforms is

$$Q(k) = 1 - \sum_{j=0}^{k-1} \binom{m}{j} \frac{\binom{i-m+j}{m-j}}{\binom{i}{m}}. \tag{2.6}$$

So far we have simplified our calculations by assuming that the same isoform is never expressed more than once for a given neuron. We can straightforwardly extend our result to deal with the situation in which we allow one of the neurons in the pair to express the same isoform multiple times, as would occur if the expression of each isoform was an independent random event. The probability $Q(k)$ that a pair of neurons shares k or more isoforms out of m is then

$$Q(k) = 1 - \sum_{j=0}^{k-1} \binom{m}{j} \frac{\binom{i-m+j}{m-j}}{\binom{i}{m}} \tag{2.7}$$

$$= 1 - \sum_{j=0}^{k-1} \binom{m}{j} \frac{\binom{i-1}{m-j}}{\binom{i+m-1}{m}} \tag{2.8}$$

In expression 2.7, the sum is over the number of ways the isoforms in common can be chosen (without repeats), times the number of ways the isoforms not in common can be chosen, divided by the total number of ways there are to choose these isoforms. Equation 2.8 simplifies the result using the identity $\binom{p}{q} = \binom{p+q-1}{q}$. As a simple numerical example, consider a neuron that has $m = 5$ isoforms, choosing from a possible $i = 40$ isoforms. The probability $Q(k)$ of having k or more isoforms in common for this case is given in Table 3.

We evaluated equation 2.8 for the same cases investigated by Hattori et al. (2009) shown in Table 1. Our corresponding results are shown in Table 4, and it can be seen that they match very closely with those in Table 1.

Table 3: A Simple Example.

Number of isoforms in common (k)	1	2	3	4	5
Probability ($Q(k)$)	0.47	0.0912	0.007	0.00018	9.21e-07

Notes: This shows how the probability $Q(k)$ that a pair of neurons has k or more isoforms in common changes with increasing k calculated using equation 2.8. Each neuron expresses $m = 5$ isoforms randomly selected from a set of $i = 40$ isoforms. It can be seen that $Q(k)$ drops very quickly with k .

The only significant differences occur when both tables give probabilities of order $< 10^{-9}$. This is expected, as Hattori et al. (2009) calculated their numerical values using a Monte Carlo sampling with 10^9 iterations, which means that the probabilities for events with probability $P \lesssim 10^{-8}$ are not expected to be sampled accurately.

2.1 Nondistinct Isoforms on the First Neuron. In the derivation above, equation 2.8 takes into account the possibility that one neuron in the pair may express the same isoform more than once. However, we retain the simplifying assumption that the first neuron chose distinct isoforms. This allows us to subtract the number of isoforms in common from the pool of total isoforms in the second term of the numerator. Obviously each neuron is equally likely to have a repeating isoform. How much does taking this into account change the probabilities? The probability that all the isoforms are distinct is given by the number of ways of choosing without repeats, divided by the number of ways of choosing with repeats. Some numerical examples are given in Table 5.

The probability that all the isoforms chosen by the first neuron are distinct approaches 1 as $i \gg m$. This means that it is highly probable that the isoforms on the first neuron will be distinct, and thus equation 2.8 holds as a good approximation. Furthermore, for low values of k , the approximation is accurate even when this assumption does not hold, because the nondistinct isoforms do not affect the result significantly. The probability of a single neuron expressing the same isoform r times (where $r = 1$ is the case where the isoforms are all distinct) is given by

$$\frac{\binom{i}{m-(r-1)} \binom{m-(r-1)}{r-1}}{\left(\binom{i}{m-(r-1)}\right) \binom{i+(r-2)}{r-1}}. \quad (2.9)$$

The first expression in this equation is the number of ways the distinct isoforms not in common can be chosen, that is, the number of ways they can be chosen without repeats, divided by the number of ways they can be chosen with repeats. The second term is the number of ways the isoforms in common can be chosen so that they match one of the isoforms already chosen, divided by the number of ways they could be chosen without

Table 4: Probability That a Pair of Neurons, Each Expressing 30 Isoforms Randomly Selected from a Set of i Isoforms, Has k or More Isoforms in Common Calculated Using Equation 2.8.

k	Number of Isoforms (i)									
	20,000	10,000	5000	2000	1000	500	200	100		
1	0.044	0.0861	0.165	0.362	0.594	0.835	0.989	1		
2	0.000919	0.00357	0.0135	0.0712	0.217	0.52	0.932	0.999		
3	1.2e-05	9.29e-05	0.000698	0.00905	0.0534	0.238	0.791	0.991		
4	1.09e-07	1.69e-06	2.53e-05	0.000816	0.0095	0.0818	0.578	0.963		
5	7.37e-10	2.28e-08	6.84e-07	5.5e-05	0.00127	0.0217	0.353	0.892		
6	3.84e-12	2.38e-10	1.43e-08	2.87e-06	0.000133	0.00451	0.178	0.763		
7	1.58e-14	1.96e-12	2.35e-10	1.18e-07	1.1e-05	0.000751	0.074	0.584		
8	5.23e-17	1.3e-14	3.12e-12	3.93e-09	7.36e-07	0.000101	0.0253	0.39		
9	1.41e-19	6.98e-17	3.36e-14	1.06e-10	4e-08	1.11e-05	0.00717	0.223		
10	3.1e-22	3.08e-19	2.97e-16	2.36e-12	1.79e-09	1.01e-06	0.00168	0.108		
11	5.65e-25	1.12e-21	2.17e-18	4.32e-14	6.6e-11	7.54e-08	0.000326	0.0443		
12	8.5e-28	3.38e-24	1.31e-20	6.55e-16	2.01e-12	4.66e-09	5.24e-05	0.0151		
13	1.06e-30	8.44e-27	6.53e-23	8.22e-18	5.08e-14	2.39e-10	6.99e-06	0.00431		
14	1.09e-33	1.74e-29	2.7e-25	8.54e-20	1.06e-15	1.01e-11	7.72e-07	0.00102		
15	9.35e-37	2.98e-32	9.25e-28	7.33e-22	1.84e-17	3.55e-13	7.04e-08	0.000198		
16	6.58e-40	4.2e-35	2.61e-30	5.19e-24	2.62e-19	1.02e-14	5.27e-09	3.16e-05		
17	3.79e-43	4.84e-38	6.03e-33	3.01e-26	3.06e-21	2.42e-16	3.23e-10	4.12e-06		
18	1.78e-46	4.55e-41	1.14e-35	1.42e-28	2.9e-23	4.64e-18	1.6e-11	4.35e-07		
19	6.75e-50	3.45e-44	1.72e-38	5.41e-31	2.22e-25	7.19e-20	6.41e-13	3.67e-08		
20	2.04e-53	2.09e-47	2.09e-41	1.64e-33	1.36e-27	8.87e-22	2.04e-14	2.46e-09		
21	4.87e-57	9.97e-51	1.99e-44	3.93e-36	6.52e-30	8.6e-24	5.08e-16	1.28e-10		

Notes: These calculations make the simplifying assumption that one neuron of the pair expresses 30 distinct isoforms. All probabilities are expressed to three significant figures.

Table 5: Probability That All Isoforms Expressed by a Neuron Are Distinct (Equation 2.9) with m Isoforms Expressed per Neuron and i Possible Isoforms.

		Number of Isoforms (i)								
		1000	1500	2000	2500	3000	3500	4000	4500	5000
m	10	0.914	0.942	0.956	0.965	0.97	0.975	0.978	0.98	0.982
	15	0.811	0.869	0.9	0.919	0.932	0.942	0.949	0.954	0.959
	20	0.684	0.776	0.827	0.859	0.881	0.897	0.909	0.919	0.927
	25	0.549	0.67	0.741	0.787	0.819	0.842	0.861	0.875	0.887
	30	0.419	0.56	0.647	0.706	0.748	0.78	0.805	0.824	0.84

Note: The probability that all isoforms are distinct approaches 1 as $i \rightarrow \infty$.

Table 6: Probability That a Neuron Expresses the Same Isoform r Times When It Expresses $m = 30$ Isoforms Randomly Selected from a Set of $i = 5000$ Isoforms.

Isoforms (r) in common	1	2	3	4	5	6	7	8
Probability	0.84	0.00493	2.6e-05	1.22e-07	5.03e-10	1.81e-12	5.54e-15	1.42e-17

Notes: The case $r = 1$ is the case where all expressed isoforms are unique. It can be seen that this case is the most likely.

restrictions. The values generated by this equation get very small quite quickly as r increases; the case where all the isoforms are distinct is the most likely (see Table 6). Thus, for the parameter values of relevance to Dscam1 expression, equation 2.8 is a very good approximation to the case where the isoforms on the first neuron can be assumed to be nondistinct. We examine the full solution without this assumption later.

3 Limits and Scale Laws

One advantage an analytic solution provides over a simulation is that it is more amenable to analysis of how the solution scales with changes in input parameters and what occurs as parameters are taken to their limits. Here, we further simplify equation 2.8 and examine how the probability of neurons being unique changes as the total number of isoforms i and the number of isoforms per neuron m are modified.

The first simplification we make is to consider only the probability $1 - Q(1)$ that a pair of neurons does not share any isoforms:

$$p = \frac{\binom{i-1}{m}}{\binom{i+m-1}{m}} \tag{3.1}$$

$$= \prod_{n=1}^m \frac{i-n}{i+n-1}, \tag{3.2}$$

where we have used the identity that $\binom{m}{0} = 1$. In this simplification we cannot examine how p scales with the degree of sharing allowed.

Next, equation 3.2 can be approximated by examining $\log p$, which gives a sum rather than a product. By approximating this sum with an integral, we find that

$$p \approx \frac{(i - 1)^{(i-1)} i^{(i+2)}}{(i - m)^{(i-m)} (i + m - 1)^{(i+m+1)}}. \tag{3.3}$$

Testing this approximation numerically, it agrees with the full solution to within 3% for $m = 30$ over the range of values of i considered in Table 7. This approximation is also extremely fast to calculate. It may be useful if probabilities for very large i , or a large number of probabilities, need to be calculated.

We can use this approximation to examine the limits and scaling for p with i and m . The limits for p as i is taken to extremes are as expected:

$$\lim_{i \rightarrow 0^+} p = 0, \tag{3.4}$$

$$\lim_{i \rightarrow \infty} p = 1. \tag{3.5}$$

That is, if there are few isoforms from which to choose, neurons are likely to share isoforms, while as the number of isoforms is increased, the probability that a pair of neurons does not share any isoforms approaches 1. In addition, we find that for large i , p asymptotically approaches 1:

$$p = 1 - \mathcal{O}\left(\frac{1}{i}\right) \text{ as } i \rightarrow \infty. \tag{3.6}$$

This means that the asymptotic approach to $p = 1$ slows as i increases.

The limits of p as m is taken to extremes is more complex because the limit as $m \rightarrow 0$ leaves a dependence on i . However, we find that

$$\lim_{m \rightarrow \infty} p = 0, \tag{3.7}$$

$$\lim_{i \rightarrow 0^+} \left[\lim_{m \rightarrow 0^+} p \right] = 0, \tag{3.8}$$

$$\lim_{i \rightarrow \infty} \left[\lim_{m \rightarrow 0^+} p \right] = 1, \tag{3.9}$$

$$\lim_{i \rightarrow 0^+} \left[\lim_{m \rightarrow i^-} p \right] = 0, \tag{3.10}$$

$$\lim_{i \rightarrow \infty} \left[\lim_{m \rightarrow i^-} p \right] = 0. \tag{3.11}$$

Table 7: Probability That a Pair of Neurons, Each Expressing 30 Isoforms Randomly Selected from a Set of i Isoforms, Has k or More Isoforms in Common Calculated Using Equation 4.2.

k	Number of Isoforms (i)									
	20,000	10,000	5000	2000	1000	500	200	100		
1	0.044	0.0861	0.165	0.362	0.594	0.835	0.989	1		
2	0.000919	0.00357	0.0135	0.0712	0.217	0.519	0.93	0.998		
3	1.2e-05	9.29e-05	0.000698	0.00905	0.0333	0.237	0.786	0.988		
4	1.09e-07	1.69e-06	2.53e-05	0.000815	0.00948	0.0814	0.571	0.954		
5	7.37e-10	2.28e-08	6.84e-07	5.49e-05	0.00127	0.0215	0.346	0.872		
6	3.84e-12	2.38e-10	1.43e-08	2.86e-06	0.000133	0.00448	0.173	0.731		
7	1.58e-14	1.96e-12	2.35e-10	1.18e-07	1.1e-05	0.000745	0.0713	0.544		
8	5.23e-17	1.3e-14	3.11e-12	3.93e-09	7.34e-07	0.0001	0.0242	0.352		
9	1.41e-19	6.98e-17	3.36e-14	1.06e-10	3.99e-08	1.1e-05	0.00679	0.194		
10	3.1e-22	3.08e-19	2.97e-16	2.36e-12	1.78e-09	9.97e-07	0.00158	0.0909		
11	5.65e-25	1.12e-21	2.17e-18	4.32e-14	6.57e-11	7.44e-08	0.000303	0.0357		
12	8.5e-28	3.38e-24	1.31e-20	6.54e-16	2e-12	4.59e-09	4.83e-05	0.0117		
13	1.06e-30	8.44e-27	6.53e-23	8.21e-18	5.06e-14	2.35e-10	6.39e-06	0.00321		
14	1.09e-33	1.74e-29	2.7e-25	8.53e-20	1.06e-15	9.95e-12	7e-07	0.000729		
15	9.35e-37	2.98e-32	9.25e-28	5.18e-24	1.83e-17	3.48e-13	6.32e-08	0.000137		
16	6.58e-40	4.2e-35	2.61e-30	5.18e-24	2.61e-19	1e-14	4.7e-09	2.1e-05		
17	3.79e-43	4.84e-38	6.03e-33	3.01e-26	3.04e-21	2.37e-16	2.85e-10	2.64e-06		
18	1.78e-46	4.55e-41	1.13e-35	1.42e-28	2.88e-23	4.54e-18	1.41e-11	2.68e-07		
19	6.75e-50	3.45e-44	1.72e-38	5.4e-31	2.21e-25	7.02e-20	5.58e-13	2.18e-08		
20	2.04e-53	2.09e-47	2.09e-41	1.64e-33	1.35e-27	8.65e-22	1.76e-14	1.41e-09		
21	4.87e-57	9.97e-51	1.99e-44	3.93e-36	6.48e-30	8.38e-24	4.35e-16	7.11e-11		

Note: All probabilities are expressed to three significant figures.

Thus, as expected, as neurons express fewer isoforms, they are less likely to share isoforms, whereas the probability of sharing an isoform rises as $m \rightarrow i$. Furthermore, as $m \rightarrow i^-$, p goes to zero asymptotically much more rapidly than $p \rightarrow 1$ with increasing i ,

$$p = \mathcal{O}(\exp(-m)) \text{ as } m \rightarrow i^- \tag{3.12}$$

Aside from demonstrating that these approximations behave correctly in the limiting case and that scale laws can be derived, these results show that the asymptotic behavior as $i \rightarrow \infty$ is much slower than as $m \rightarrow i^-$. This means that for large i and m , increasing m will drive $p \rightarrow 0$ rapidly, whereas maintaining the same value of p would require a much larger increase in i . Therefore, if neurons could tolerate only little isoform sharing, one would expect to find low values of m rather than extremely large i , because i would have to be unrealistically large in order to achieve the same low values of p .

4 An Exact Solution Based on Collision Probabilities _____

A recent result from the combinatorics literature solves the Dscam1 isoform problem directly without requiring any approximations to be made (Nakata, 2008). Consider the occupancy problem, where m_1 black balls and m_2 white balls are thrown into i bins. Imagine that white balls represent the isoforms from one neuron and black balls the isoforms expressed by another neuron. Throwing the balls into i bins is analogous to choosing the type of each isoform from i possible isoforms. A “collision” occurs if a bin contains both white and black balls. This is analogous to an isoform being shared between two neurons, where a black and a white ball that both land in the same bin is the same as both neurons choosing the same isoform. In this calculation, there are no restrictions that either neuron chooses distinct isoforms, and this problem is thus exactly analogous to the Dscam1 isoform problem. The probability that the two neurons have k isoforms in common is the same as the probability that k bins have both white and black balls. The general formula derived by Nakata (2008) describing this case is

$$C(j) = \frac{1}{i^{m_1+m_2}} \sum_{a=j}^{m_1} \sum_{b=j}^{m_2} \left\{ \begin{matrix} m_1 \\ a \end{matrix} \right\} \left\{ \begin{matrix} m_2 \\ b \end{matrix} \right\} \frac{(a)_{m_1} (b)_{m_2} (i)_{a+b+j}}{j!} \tag{4.1}$$

Here $C(j)$ is the probability of exactly j collisions, $\left\{ \begin{matrix} m \\ a \end{matrix} \right\}$ is the Stirling number of the second kind (Graham, Knuth, & Patashnik, 1994), which gives the number of ways of partitioning m objects into a nonempty subsets, and $(a)_j = \binom{a}{j} j! = a! / (a - j)!$ Here m_1 is the number of isoforms on the first neuron, and m_2 is the number of isoforms on the second neuron (our previous

results have taken into account only the case where $m_1 = m_2$). Thus, the probability $Q(k)$ of k or more collisions is

$$Q(k) = 1 - \sum_{j=0}^{k-1} C(j). \quad (4.2)$$

Probabilities $Q(k)$ for the cases addressed in Tables 1 and 4 are tabulated in Table 7 with this exact solution. It can be seen that the only significant change from the approximate solution (see table 4 and equation 2.8) occurs when there is a very low number of isoforms from which to choose.

This formula can be used to calculate $Q(k)$ without the assumption of equation 2.8 that one neuron chooses all distinct isoforms. However, the proof of equation 4.2 is difficult to understand intuitively. Equation 2.8 is accurate over biologically relevant parameter ranges and is amenable for use in examining how the probabilities scale with parameter changes. It is also faster to evaluate numerically due to a reduced number of terms and can be accurately calculated using lower-precision numbers.

However, care must still be taken when evaluating both equations 2.8 and 4.2 because ratios of large factorials have to be calculated and summed. An arbitrary precision arithmetic library should be used to avoid intermediate rounding errors caused by a lack of precision. These errors can seriously affect the result as they accumulate in the summing process. We performed all calculations and plots using an arbitrary precision Python library (Stein et al., 2010). Our code is available on request. Calculations for a single probability take seconds on a modern computer (Intel Core 2 Duo, 2.8 Ghz).

5 Population of Neurons

Having considered the probability of a neuron pair expressing distinct isoforms, we now consider the probability that a set of n neurons all correctly self-avoids. With n neurons, there are $N = \frac{n(n-1)}{2}$ possible neuron pairs. What is the probability P that all pairs are distinct, that is, share fewer than k isoforms? This is just the probability that each pair is distinct to the power of the number of pairs: $P = (1 - Q(k))^N$. Solving for n gives

$$n = \frac{1 + \sqrt{1 + \frac{8 \log P}{\log(1-Q(k))}}}{2}. \quad (5.1)$$

Therefore, given the number of isoforms allowed to be in common, the number of total isoforms, and the isoforms per neuron, the maximum number of neurons that are distinct at a given probability $Q(k)$ can be calculated. Some examples are shown in Figure 2A for $P = 95\%$.

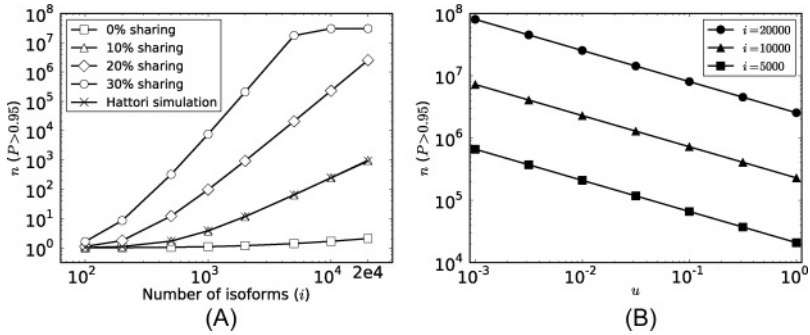


Figure 2: (A) The maximum number of distinct neurons n that can be created while maintaining a 95% probability that all neuron pairs are distinct as a function of the total number of isoforms i as calculated by equation 5.1. We show the results when there are $m = 30$ isoforms per neuron and when varying degrees of sharing are allowed before neurons are no longer considered distinct. Both axes are log scale. In the wild type animal, $i \approx 20,000$ ($2e4$). The Hattori simulation line was plotted using values from Hattori et al. (2009) with 10% sharing and $m = 30$. It is clear that these values are consistent with the values generated by equation 5.1. (B) The number of distinct neurons that can be created with a probability of 95% that all contacting neurons will have a distinct identity when neurons contact with a likelihood u . This is calculated for several values of i with $m = 30$ isoforms per neuron and 20% sharing allowed.

However, this gives only a lower bound for n , as it requires all the possible neuron pairs to be distinct. Biologically this may not be a necessary requirement because not all neurons will come into contact. To allow for this, we introduce an extra term u , which is the probability that a pair of neurons comes into contact. We then have

$$n = \frac{1 + \sqrt{1 + \frac{8 \log P}{u \log(1-Q(k))}}}{2}. \tag{5.2}$$

If u is small, it can have a significant affect on the number of distinct neurons that can be created with a limited number of isoforms (see Figure 2B).

An important special case to consider is to assume that each neuron encounters a fixed number c other neurons, regardless of the total number of neurons. This occurs in many geometrical arrangements. In this case, $u = c/n$, so we find that

$$P = (1 - Q)^{\frac{c}{2}(n-1)}, \tag{5.3}$$

$$n = \frac{2}{c} \frac{\log p}{\log(1 - Q)} + 1. \tag{5.4}$$

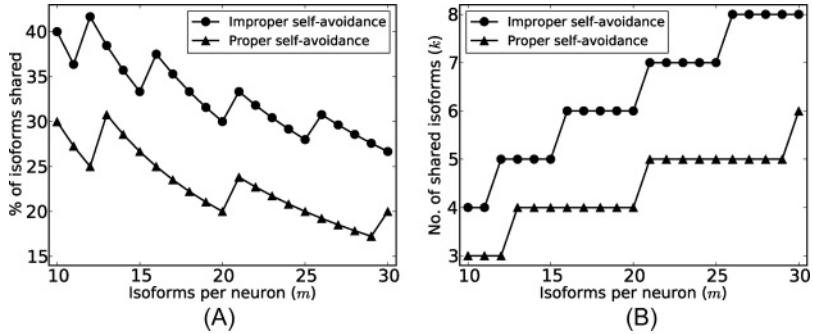


Figure 3: The minimum percentage (A) and absolute number (B) of isoform sharing that would need to be tolerated in order to maintain a 95% probability that all mushroom body neurons are distinct. We show the degree of sharing required for the situation in which Hattori et al. (2009) found proper self-avoidance ($i = 4752$ total isoforms) and for the situation in which they found improper self-avoidance ($i = 1152$ total isoforms). Bumps and nonmonotonicity in the graph are due to the number of isoforms that are shared, changing discretely as m is increased.

Note that in this case, increasing n still requires decreasing Q , but is now of order $\mathcal{O}(1/(1 - Q))$ rather than $\mathcal{O}(1/\sqrt{1 - Q})$.

5.1 Mushroom Body Neurons. A specific case of biological interest is mushroom body (MB) development in *Drosophila*. Most MB neurons consist of two branches that split into two paths and form the two lobes of the MB. Without *Dscam1*, only one lobe is formed (Hattori et al., 2009), which suggests that *Dscam1*-mediated self-recognition and repulsion play an important role in the segregation of the two branches. The MB consists of 2500 neurons, yielding a requirement of $Q(k) = 1.64 \times 10^{-8}$ for correct self-avoidance if a 95% probability of all pairs being unique is acceptable. MB neurons have been estimated to express 10 to 30 isoforms each. By examining mutants with reduced *Dscam1* diversity, Hattori et al. (2009) were able to show that expressing only 4752 isoforms rather than all possible 19,008 isoforms was sufficient for appropriate branching, but that reducing this number further to 1152 isoforms was not sufficient.

Using these numbers, we investigated the degree of sharing that would need to be tolerated in order to maintain proper self-avoidance for varying values of m , the number of isoforms expressed by each neuron, and i the total number of isoforms (see Figure 3A). We considered both the 4752 isoform case where Hattori et al. (2009) found appropriate branching occurred, and the 1152 isoform case where branching was impeded. These two situations may provide a bound on the degree of isoform sharing before neurons are no longer considered distinct. We found that even for 30 isoforms per

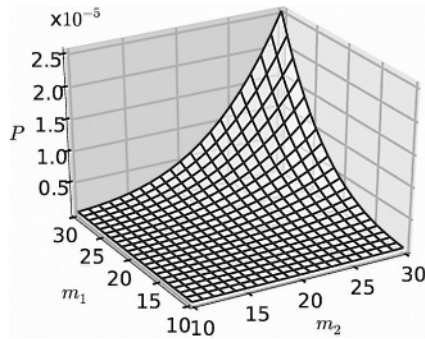


Figure 4: The relationship between the number of isoforms expressed on each neuron m_1 and m_2 and the probability that the pair of neurons has one or more isoforms in common. Calculated with $i = 5000$ isoforms in total.

neuron, at least 25% sharing is required for proper self-avoidance when $i = 4725$ isoforms.

5.2 Absolute Rather Than Relative Sharing. The above calculations suggest that quite a high proportion of sharing must be allowed for proper self-avoidance. One important consideration is that it may be that repulsion is triggered by an absolute number of homophilic isoform bindings, independent of the total number of isoforms each neuron expresses. This is explored in Figure 3B, which shows an approximately linear increase in the absolute number of isoforms that must be shared as the number of isoforms per neuron increases.

Furthermore, if an absolute number of isoforms needs to be shared to trigger repulsion, then it would not matter how many isoforms each neuron expresses as long as they shared over a certain number of isoforms. This would imply that the number of isoforms per neuron is irrelevant when determining if a set of isoforms is unique.

5.3 Variable Numbers of Isoforms on Each Neuron. So far we have assumed, following Hattori et al. (2009), that the same number of isoforms is expressed on each neuron. However, biologically this assumption is unlikely to be the case. Does allowing this number to be chosen stochastically from some distribution independently for each neuron change the basic conclusions? To explore this, we used equation 4.2 to calculate how the probability of having a number of isoforms in common changes with the number of isoforms expressed by each neuron (see Figure 4). It can be seen that the effect of a mismatch between the number of isoforms expressed by one neuron versus another depends strongly on the absolute numbers involved. We also calculated the probability distribution for the difference

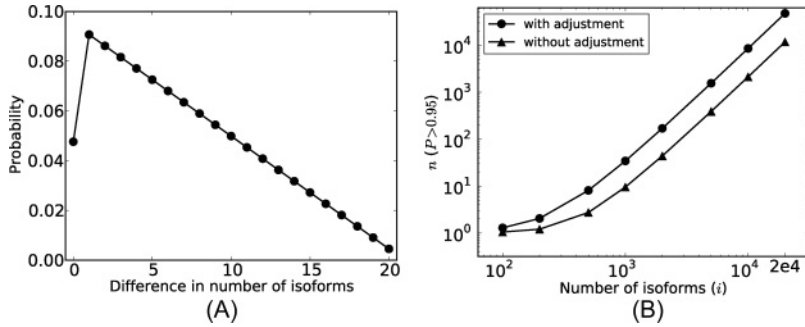


Figure 5: (A) The probability distribution of the difference in the number of isoforms expressed between two neurons if each neuron expresses a number of isoforms m chosen independently from a discrete uniform distribution with a range $[10, 30]$. It is much more likely that a randomly chosen pair of neurons differs in the number of isoforms they express than that they do not. (B) The number of possible distinct neurons with a probability of 95% that the neurons are unique when each neuron expresses exactly $m = 30$ isoforms (unadjusted, triangles) and when the number of isoforms expressed by each neuron is chosen stochastically (adjusted, circles). In these calculations, neurons are allowed to share up to $k = 4$ isoforms while still being considered distinct. In the wild type animal, $i \approx 20000$ ($2e4$).

in number of isoforms between two neurons if each neuron expresses a random number of isoforms chosen from a uniform distribution between 10 to 30 (see Figure 5A). Figure 5B shows how the number of neurons n that can be created while maintaining correct repulsion can increase when the number of isoforms per neuron is allowed to vary. This adjustment simply scales the earlier results and does not significantly change the overall conclusions.

6 Discussion

Several recent experiments have dramatically increased our understanding of the role played by *Dscam1* in *Drosophila* neural development (reviewed in Millard & Zipursky, 2008; Zipursky & Sanes, 2010). In particular, it is now known that *Dscam1* plays a key role in neuronal self-avoidance (Zhan et al., 2004; Matthews et al., 2007) and that a large number of distinct isoforms are necessary for normal development (Hattori et al., 2008; Chen et al., 2006). Hattori et al. (2009) began a mathematical investigation of these phenomena by using Monte Carlo simulations to estimate the probabilities involved in *Dscam1* self-recognition. We have expanded on their work by demonstrating that it is possible to calculate a closed-form analytic solution to this problem, as well as a much simpler but still accurate closed-form

approximation. The advantage offered by these analytic solutions is that they lend greater clarity to the underlying combinatorial issues and provide a basis for analysis of how biologically relevant probabilities scale as key parameters vary. Although the mathematics cannot answer experimental questions, it can place constraints on the possible mechanisms involved in Dscam1 mediated self-avoidance and help inform future experiments.

6.1 What Must Neuronal Self-Avoidance Achieve? Neuronal self-avoidance is necessary to ensure that processes from the same neuron target only processes from other neurons. For successful self-avoidance, neurons must both correctly detect when they encounter processes from the same neuron (avoiding false negatives) and when they encounter processes from other neurons (avoiding false positives). The difficulty of this problem increases as neuronal numbers and density are increased and as the tolerable fraction of detection errors is decreased.

In addition, it is likely that some small amount of both false positives and false negatives in self-avoidance does not result in measurable developmental defects. There are currently few results that provide insight into the level of acceptable error that might exist, and the issue is complicated by the likelihood of compensatory mechanisms, such as redundancy in synapse formation and activity-driven synaptic refinement. Following Hattori et al. (2009), we have allowed a 5% false-positive rate when concrete numbers were required. Our model does not explicitly include false negatives, but we consider this in more detail in the next section.

We have used the parameter n to denote the number of neurons that must self-avoid. However, interpreting this parameter as the total number of neurons in the brain is obviously overly pessimistic. The geometry of growing neuronal processes means that two randomly chosen neurons are unlikely to encounter one another; we denoted the likelihood of encounter by u . As described earlier, one possibility is that neurons encounter a fixed number of other neurons, regardless of the total number of neurons. In this case, we found that the difficulty of correct self-avoidance (and thus the need for increasing the total number of isoforms i) increases only slowly as more neurons are introduced. This reinforces the idea that the number of isoforms required for correct self-avoidance is likely to be affected more dramatically by changes in neuronal density, growth patterns, and geometry than by adding more neurons. This scaling may be important in allowing evolutionary increases in brain size without requiring fundamental changes in self-avoidance mechanisms.

6.2 Autapses and False Negatives. Throughout this work, we have assumed that autapses, synapses between processes belonging to the same neuron, are not beneficial. However, autapses do occur *in vivo*, and recent work has demonstrated some functional uses for them (Bacci, Huguenard, & Prince, 2003; Bekkers, 2003, 2009; Saada, Miller, Hurwitz, & Susswein,

2009). Most of this work has been in mammalian brains, but it has been shown that autapses play a functional role in *Aplysia* (Saada et al., 2009). Therefore, it is plausible that in some *Drosophila* brain regions, a certain level of autapse formation may be necessary for function. However, since *Dscam1* knockouts display serious developmental defects and *Dscam1* is known to mediate self-avoidance, this demonstrates that an abundance of autapse formation is harmful. Further, it demonstrates that synaptic refinement of dysfunctional autapses during development is not enough; functioning self-avoidance mechanisms are needed for normal development to occur.

Little is known about autapses in *Drosophila*; however, their potential presence does not significantly affect our findings. From the point of view of modeling self-avoidance, autapses represent a false negative, and our model does not explicitly include false negatives. One way they could be included would be to suppose that *Dscam1* mediated self-avoidance is stochastic, so that autapses occasionally form during self-encounters. This could be represented as an additional interpretation of the parameter u to include the probability that two processes expressing the same *Dscam1* isoforms will repulse. Since most synapses that are formed are not autapses, the change in u needed to model the development of autapses would be small. As shown in Figure 2B, u acts as a scaling parameter, and small changes in u do not significantly affect the combinatorial problems we have considered.

6.3 What Defines the Uniqueness of a Neuron? An important outstanding biological question in self-avoidance is: what determines the uniqueness of a set of isoforms? If contact-dependent repulsion occurs only when all the isoforms are identical, the important probability is that of having exactly m isoforms in common between two neurons. This would allow 10^{27} to 10^{35} neurons with distinct identities (using 1152 to 4752 isoforms), which is many more than are present in the brain of *Drosophila*. However, it is unlikely that the mechanism that allows contact-dependent binding and repulsion requires an identical match to every isoform. On the other hand, if the presence of a single shared isoform is enough to promote avoidance, then few unique neurons are likely to be produced, and a large number of false positives will occur. Hence, there must be a threshold for the number of shared isoforms, whether in total number or as a percentage of the total, such that greater than this number of isoforms promotes avoidance.

The fact that neurons express multiple *Dscam1* isoforms provides evidence that some isoform sharing is acceptable before repulsion occurs: if a single shared isoform is all that is required to generate repulsion, it would be optimal for each neuron to express only a single isoform, which would ensure self-avoidance while minimizing false positives. If some degree of sharing is allowed, there is then a trade-off between expressing fewer isoforms per neuron, and thus increasing the probability that a pair of neurons does not share any isoforms, versus allowing a greater degree of sharing,

but increasing the number of isoforms that each neuron expresses. One advantage that expressing multiple isoforms provides is that the redundancy could ensure that self-avoidance is robust even if some isoforms do not bind correctly. The trade-offs involved in *Dscam1* isoform expression to minimize both metabolic cost and self-avoidance errors could be an interesting avenue for future work.

One potential experimental avenue for testing hypotheses about how *Dscam1* isoforms determine the uniqueness of a neuron is through single cell mutation. Here all the neurons are mutated to express the same *Dscam1* isoform set, while a single cell expresses a different set. Single cell experiments have previously been used to demonstrate that different isoform complements repel (Matthews et al., 2007); it may be possible to use such experiments to measure the degree of sharing allowed before repulsion occurs. There is evidence that although *Dscam1* isoforms are largely homophilic, some degree of heterophilic binding also occurs (Wojtowicz et al., 2007). This heterophilic binding is often weaker and usually occurs only with a handful of other isoforms. As such, the existence of such binding has little effect on our results. It could be modeled by simply reducing the total number of isoforms i by some amount, possibly even a fractional number (to account for weak binding).

6.4 Deterministic Expression of Isoforms. Our calculations have been based on the assumption that the expression of isoforms on a given neuron is an entirely stochastic event. Another possibility is that neurons possess information about which other neurons in the population they are more likely to encounter during development. This information could be used to ensure that neurons that are likely to encounter one another express different isoforms to minimize self-avoidance false positives, while dramatically lowering the number of distinct isoform identities required for correct self-avoidance. This situation could be formalized as a random graph process with the probability of a neuron encountering another neuron represented by the probability of a vertex existing between the two neurons.

One possible biological implementation of this idea would be that neurons near one another, and therefore maybe more likely to encounter one another in the future, use intercellular signaling to ensure that nearby cells express isoforms that are distinct. Another possibility is that immunological mechanisms may eliminate cells that are nearby, but expressing similar isoform sets, in order to minimize erroneous avoidance during later development. While too little is known about the biological mechanisms of *Dscam1* isoform selection to constrain a random graph theory model of self-avoidance, this approach may be useful for interpreting future findings.

6.5 Self-Avoidance in Vertebrates. Functioning self-avoidance mechanisms are essential for normal development in *Drosophila*, and it seems likely that such a fundamental organizing principle could also be at work in

developing vertebrate brains. Dscam in vertebrates has been shown to undergo homophilic binding (Agarwala, Nakamura, Tsutsumi, & Yamakawa, 2000), and mice *Dscam* knockouts have neural defects (Fuerst, Koizumi, Masland, & Burgess, 2008). However, in contrast to invertebrate *Dscam1*, *Dscam* in vertebrates does not have a large number of splice variants, which are essential for correct self-avoidance (Yamakawa et al., 1998). Therefore, *Dscam1* is unlikely to act as a cell surface recognition molecule in vertebrates (Hattori et al., 2008) and may play a role more similar to *Dscam2* in *Drosophila*: establishing mosaic patterns that require only a small number of isoforms.

However, another set of genes in vertebrates, the clustered protocadherins (*Pcdh*), seems to share some of the properties of *Drosophila* *Dscam1* (Morishita & Yagi, 2007), which make them a promising candidate for self-avoidance molecules. The *Pcdh* gene encodes three types of protein: *Pcdh- α* , *Pcdh- β* , and *Pcdh- γ* , each containing 14, 22, and 22 isoforms, respectively. *Pcdh- α* and *- γ* clusters have a constant cytoplasmic domain with a variable ectodomain. *Pcdh- β* has both a variable cytoplasmic domain and a variable ectodomain. The diversity in the *Pcdh* isoforms is through separate promoters that alter gene expression in each cell rather than due to splice variants (Zipursky & Sanes, 2010). Each variable exon is expressed monoallelically, with multiple variable exons expressed in each cell from both alleles (Esumi et al., 2005).

Recently Schreiner and Weiner (2010) showed that *Pcdh- γ* exhibits isoform-specific homophilic binding. They tested 7 of the 22 *Pcdh- γ* isoforms, and each isoform showed homophilic binding activity. This is similar to the isoform-specific homophilic binding that *Dscam1* exhibits. However, there are some important differences in *Pcdh* expression, indicating that it may act differently from *Dscam1* in cellular identification. *Pcdh* has much less diversity compared with *Dscam1* (~60 isoforms compared with ~20,000 isoforms), and the number of distinct isoforms expressed per neuron is significantly lower for *Pcdh*; 3 to 6 isoforms per neuron for *Pcdh* compared with 10 to 30 for *Dscam1* (Zipursky & Sanes, 2010). Monoallelic expression of *Pcdh* restricts the number of isoforms expressed per neuron (Esumi et al., 2005). It has been proposed that each neuron expresses only two *Pcdh- α* and two *Pcdh- γ* (Kaneko et al., 2006). This constraint on expression would further restrict the number of combinations of isoforms available for self-avoidance. It would be possible to create a reasonable number of distinct neurons under these conditions, but only if the presence of a single different isoform could distinguish one set of isoforms from another. This would produce approximately 10^2 neurons, which may be sufficient depending on the geometry of neural development.

One possible model for increasing the diversity of *Pcdh* interactions is the formation of tetramers. Previously it has been shown that *Pcdh- γ* forms tetramers with other *Pcdh- γ* isoforms on the same neuron and it is highly probable that these tetramers can also include *Pcdh- α* (Schreiner & Weiner,

2010). It is possible that the composition of Pcdh tetramers determines the specificity of the interactions with other neurons. If tetramer binding could occur only between two rotationally identical tetramers, then this would almost certainly provide the diversity needed for Pcdh to act as a cell-recognition molecule. Almost 250,000 tetramers can be formed from Pcdh- γ alone (i.e., 22^4), and almost 60,000 unique tetramers can be formed, taking into account the rotation of the tetramer in the membrane. There are six ways that four distinct isoforms could be arranged in a tetramer, which more than doubles the number of distinct neurons that can be formed. Tetramer formation within the cell and subsequent translocation to the cell surface would be an ideal mechanism for providing each neuron with a unique identity; otherwise, some deterministic mechanism would be needed to ensure that the same tetramer forms at all locations on the cell surface.

It still remains unclear whether Pcdh plays a role in neuronal patterning because currently no link exists between Pcdh diversity and function. However, since it shares several key properties with Dscam1—homophilic-specific binding and a stochastic expression of multiple isoforms—it seems a likely candidate for self-recognition during neurite development in vertebrates.

Acknowledgments

We are very grateful to Sean Millard for many helpful discussions. We also thank Massimo Hilliard, Hugh Simpson, Andrew Thompson, and Rowan Tweedale for helpful comments on earlier drafts of this letter and Larry Zipursky, Daisuke Hattori, and Lukasz Salwinski for providing further details of the stochastic simulations in Hattori et al. (2009). Funding was provided by NHMRC Project Grant 631532, Program Grant RPG0029/2008-C from the Human Frontiers Science Program, and the Queensland Brain Institute.

References

- Agarwala, K. L., Nakamura, S., Tsutsumi, Y., & Yamakawa, K. (2000). Down syndrome cell adhesion molecule DSCAM mediates homophilic intercellular adhesion. *Molecular Brain Research*, 79(1–2), 118–126.
- Bacci, A., Huguenard, J. R., & Prince, D. A. (2003). Functional autaptic neurotransmission in fast-spiking interneurons: A novel form of feedback inhibition in the neocortex. *Journal of Neuroscience*, 23(3), 859–866.
- Bekkers, J. M. (2003). Synaptic transmission: Functional autapses in the cortex. *Current Biology*, 13(11), R433–R435.
- Bekkers, J. M. (2009). Synaptic transmission: Excitatory autapses find a function? *Current Biology*, 19(7), R296–R298.
- Chen, B. E., Kondo, M., Garnier, A., Watson, F. L., Püettmann-Holgado, R., Lamar, D. R., et al. (2006). The molecular diversity of Dscam is functionally required for neuronal wiring specificity in *Drosophila*. *Cell*, 125(3), 607–620.

- Chilton, J. K. (2006). Molecular mechanisms of axon guidance. *Developmental Biology*, 292(1), 13–24.
- Dickson, B. J. (2002). Molecular mechanisms of axon guidance. *Science*, 298(5600), 1959–1964.
- Esumi, S., Kakazu, N., Taguchi, Y., Hirayama, T., Sasaki, A., Hirabayashi, T., et al. (2005). Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons. *Nature Genetics*, 37(2), 171–176.
- Fuerst, P. G., Koizumi, A., Masland, R. H., & Burgess, R. W. (2008). Neurite arborization and mosaic spacing in the mouse retina require DSCAM. *Nature*, 451(7177), 470–474.
- Graham, R. L., Knuth, D. E., & Patashnik, O. (1994). *Concrete mathematics: A foundation for computer science*. Reading, MA: Addison-Wesley.
- Hattori, D., Chen, Y., Matthews, B. J., Salwinski, L., Sabatti, C., Grueber, W. B., et al. (2009). Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. *Nature*, 461(7264), 644–648.
- Hattori, D., Millard, S. S., Wojtowicz, W. M., & Zipursky, S. L. (2008). Dscam-mediated cell recognition regulates neural circuit formation. *Annual Review of Cell and Developmental Biology*, 24, 597–620.
- Huber, A. B., Kolodkin, A. L., Ginty, D. D., & Cloutier, J.-F. (2003). Signaling at the growth cone: Ligand-receptor complexes and the control of axon growth and guidance. *Annual Review of Neuroscience*, 26, 509–563.
- Kaneko, R., Kato, H., Kawamura, Y., Esumi, S., Hirayama, T., Hirabayashi, T., et al. (2006). Allelic gene regulation of Pcdh-alpha and Pcdh-gamma clusters involving both monoallelic and biallelic expression in single Purkinje cells. *Journal of Biological Chemistry*, 281(41), 30551–30560.
- Matthews, B. J., Kim, M. E., Flanagan, J. J., Hattori, D., Clemens, J. C., Zipursky, S. L., et al. (2007). Dendrite self-avoidance is controlled by Dscam. *Cell*, 129(3), 593–604.
- Millard, S. S., & Zipursky, S. L. (2008). Dscam-mediated repulsion controls tiling & self-avoidance. *Current Opinion in Neurobiology*, 18(1), 84–89.
- Morishita, H., & Yagi, T. (2007). Protocadherin family: Diversity, structure, and function. *Current Opinion in Cell Biology*, 19(5), 584–592.
- Mortimer, D., Fothergill, T., Pujic, Z., Richards, L. J., & Goodhill, G. J. (2008). Growth cone chemotaxis. *Trends in Neurosciences*, 31(2), 90–98.
- Nakata, T. (2008). Collision probability for an occupancy problem. *Statistics and Probability Letters*, 78, 1929–1932.
- O'Donnell, M., Chance, R. K., & Bashaw, G. J. (2009). Axon growth and guidance: Receptor regulation and signal transduction. *Annual Review of Neuroscience*, 32, 383–412.
- Saada, R., Miller, N., Hurwitz, I., & Susswein, A. J. (2009). Autaptic excitation elicits persistent activity and a plateau potential in a neuron of known behavioral function. *Current Biology*, 19(6), 479–484.
- Schreiner, D., & Weiner, J. A. (2010). Combinatorial homophilic interaction between gamma-protocadherin multimers greatly expands the molecular diversity of cell adhesion. *Proceedings of the National Academy of Sciences of the United States of America*, 107(33), 14893–14898.

- Stein, W., Abbott, T., Abshoff, M., Ajanki, A., Albrecht, M., Alexander, N., et al. (2010). *Sage mathematics software (Version 2.6.0)*. Sage Development Team. <http://www.sagemath.org>.
- Wojtowicz, W. M., Wu, W., Andre, I., Qian, B., Baker, D., & Zipursky, S. L. (2007). A vast repertoire of Dscam binding specificities arises from modular interactions of variable IG domains. *Cell*, *130*(6), 1134–1145.
- Yamakawa, K., Huot, Y. K., Haendelt, M. A., Hubert, R., Chen, X. N., Lyons, G. E., et al. (1998). DSCAM: A novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system. *Human Molecular Genetics*, *7*(2), 227–237.
- Zhan, X.-L., Clemens, J. C., Neves, G., Hattori, D., Flanagan, J. J., Hummel, T., et al. (2004). Analysis of Dscam diversity in regulating axon guidance in *Drosophila* mushroom bodies. *Neuron*, *43*(5), 673–686.
- Zipursky, S. L., & Sanes, J. R. (2010). Chemoaffinity revisited: Dscams, protocadherins, and neural circuit assembly. *Cell*, *143*(3), 343–353.

Received December 15, 2010; accepted March 13, 2011.